Journal of Network and Computer Applications 34 (2011) 1732–1742

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Monitoring abnormal network traffic based on blind source separation approach ☆

Tao Qin [a,*], Xiaohong Guan [a,b], Wei Li [a], Pinghui Wang [a], Qiuzhen Huang [a]

[a] SKLMS Lab and MOE KLINNS Lab, Xi'an Jiaotong University, Xi'an 710049, China
[b] Department of Automation and TNLIST Lab, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 22 August 2010
Received in revised form
12 May 2011
Accepted 12 June 2011
Available online 24 June 2011

Keywords:
Network traffic monitoring
Abnormal behavior detection
DFlow model
Blind source separation
Scale space filter

ABSTRACT

The randomness in network behaviors poses serious challenges for discovering abnormal patterns in network traffic flows. This paper presents a systematic approach for monitoring abnormal network traffic. The DFlow model is proposed to reduce the flow records and extract four features to capture the traffic patterns. The blind source separation method is applied to obtain the routine and abnormal behaviors from those features. A scale space filter is applied to filter the randomness in the traffic flows without affecting the behavior patterns. A threshold is selected based on a systematic criterion to evaluate the degree of abnormality. The contributions of different traffic features to the abnormal behavior detection are analyzed. It is found that the number of connection degree is the most important feature for traffic monitoring. A salient feature of this method is that it is effective for detecting the abnormal behaviors not associated with significant changes in traffic volumes. Another advantage of the new method is that no supervised learning process is needed. This is very important since high quality labeled samples are very difficult to acquire in actual networks especially the data traces associated with attacks. The experimental results based on the actual network data show that the method presented in the paper is effective for monitoring abnormal traffic flows in the gigabytes traffic environment and the accuracy is above 95%.

1. Introduction

The frequent outbreaks of large-scale network security incidents, such as worms, distributed denial of service (DDOS) and other anomalous behaviors, pose serious threats to normal Internet operation, which has become the vital infrastructure in the modern information age (Staniford et al., 2002). Network monitoring is extremely important for network security. Traffic flow is one of the most important data sources for monitoring the network anomalies since the abnormal behaviors over the Internet (mal-code spreading, network resource abuse and malfunction of network equipment) would cause changes in the normal flow patterns. On the other hand, the statistical patterns of the network flow are also affected by the "noise" resulting from the random feature of network traffic. This causes serious challenges for discovering the abnormal traffic flows and thus makes it a very important topic in network security research.

In the past years, the characteristics of traffic patterns were widely studied and many machine learning methods were proposed for detecting abnormal behaviors (Zhang et al., 2005). Some of those methods are tested by the KDD datasets (KDD Cup, 1999). However in actual networks, it is very difficult to obtain high quality labeled traffic samples and it is very difficult to carry out the training process. The frequent outbreak of some unknown attacks with new traffic flow patterns also pose challenge to those methods. Therefore, it is desirable to have a method for discovering abnormal behaviors without involving training process.

There are mainly two kinds of network behaviors reflected in the traffic flow patterns: the routine behaviors and the anomalous behaviors (Lakhina et al., 2004b). The network traffic can be decomposed into alpha/beta "signals" with different frequencies and the high frequencies are usually associated with the abnormal behaviors. The routine behaviors reflect the users' habitual activities, while the abnormal traffics aim to compromise or disable hosts or networks and the characteristics of the associated behaviors are unknown due to the inherent irregularities. The signals associated with the abnormal behaviors usually follows the non-Gaussian distributions (Sarvotham et al., 2001; Lakhina et al., 2004b).

In this paper, we present an approach for extracting the abnormal behavior index from the traffic measurement features

☆The research presented in this paper is supported in part by the National Natural Science Foundation (60825202, 60921003), 863 High Tech Development Plan (2007AA01Z475, 2007AA01Z480, 2007AA01Z464, 2008AA01Z415) and 111 International Collaboration Program of China.

* Corresponding author. Fax: +86 2982664603.
E-mail address: tqin@sei.xjtu.edu.cn (T. Qin).

1084-8045/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2011.06.006